Family classification and integrative analysis of molybdenum cofactor biosynthesis proteins and related proteins

Xianying Wei Anastasia N. Nikolskaya Cathy H. Wu
Protein Information Resource
Georgetown University Medical Center
Georgetown University Medical Center, Washington, DC, USA

PIRSF is a network protein family classification system that reflects evolutionary relationships of full-length proteins. The primary PIRSF classification unit is the *homeomorphic family* whose members are both *homologous* (evolved from a common ancestor) and *homeomorphic* (sharing full-length sequence similarity and a common domain architecture). Here, we present family classification and detailed annotation of some of the molybdenum cofactor (Moco) biosynthesis proteins. PIRSF classification system allows annotation of specific biological as well as generic biochemical functions, and sequence features that are family (or subfamily) specific.

Moco is the essential component of a diverse range of redox-active enzymes. The cofactor consists of a mononuclear molybdenum coordinated by the dithiolene moeity of tricyclic molybdopterin (MPT). Genes involved in Moco biosynthesis have been identified in eubacteria, archaea and eukaryotes. Moco biosynthesis pathway can be divided into three universal stages:

1) Precursor Z biosynthesis (e.g., familiesPIRSF003315 and PIRSF004845),
2) MPT biosynthesis (e.g., families PIRSF006462, and PIRSF005983),
3) Moco biosynthesis with molybdenum incorporation (e.g., families PIRSF036627, PIRSF036598 with subfamily PIRSF500055),
4) An addditional step of Moco dinucleotide biosynthesis (e.g., families PIRSF036607 and PIRSF036622) is also required in eubacteria.

Domain shuffling is observed in Moco biosynthesis-related proteins families, and functions of some families are predicted based on their domain architectures. Protein families sharing domains with Moco biosynthesis protein also have been investigated. Interestingly, MPT binding domain PF00994 shared by most Moco molybdenum incorporation proteins is also present in competence/damage-inducible protein CinA (PIRSF006728) and some eukaryotic FAD synthetases (PIRSF036620).

Curated Moco biosynthesis protein families illustrate how various types of data from diverse sources are integrated in the PIRSF system to allow the user a faster and more accurate analysis of protein sequence and function. Such integrative analysis includes position-specific site rules (active site residues annotation), 3D structure information, genome context, phylogenetic comparison, etc.

Data integration and analysis allows standardizing nomenclature and gaining insights into the evolution of Moco biosynthesis proteins and of the corresponding pathway.